

Chapter 465

Equivalence Tests for Two Means (Simulation)

Introduction

This procedure allows you to study the power and sample size of an equivalence test comparing two means from independent groups. Schuirmann's (1987) two one-sided tests (TOST) approach is used to test equivalence. The t-test is commonly used in this situation, but other tests have been developed for use when the t-test assumptions are not met. These additional tests include the Mann-Whitney U test, Welch's unequal variance test, and trimmed versions of the t-test and the Welch test.

Measurements are made on individuals that have been randomly assigned to, or randomly chosen from, one of two groups. This *parallel-groups* design may be analyzed by a TOST equivalence test to show that the means of the two groups do not differ by more than a small amount, called the margin of equivalence.

The two-sample t-test is commonly used in this situation. When the variances of the two groups are unequal, Welch's t-test is often used. When the data are not normally distributed, the Mann-Whitney (Wilcoxon signed-ranks) U test and, less frequently, the trimmed t-test may be used.

The details of the power analysis of equivalence test using analytic techniques are presented in another *PASS* chapter and they won't be duplicated here. This chapter will only consider power analysis using computer simulation.

Technical Details

Computer simulation allows us to estimate the power and significance level that is actually achieved by a test procedure in situations that are not mathematically tractable. Computer simulation was once limited to mainframe computers. But, in recent years, as computer speeds have increased, simulation studies can be completed on desktop and laptop computers in a reasonable period of time.

The steps to a simulation study are as follows.

1. Specify how the test is carried out. This includes indicating how the test statistic is calculated and how the significance level is specified.
2. Generate random samples from the distributions specified by the alternative hypothesis. Calculate the test statistics from the simulated data and determine if the null hypothesis is accepted or rejected. Tabulate the number of rejections and use this to calculate the test's power.

Equivalence Tests for Two Means (Simulation)

3. Generate random samples from the distributions specified by the null hypothesis. Calculate each test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. Tabulate the number of rejections and use this to calculate the test's significance level.
4. Repeat steps 2 and 3 several thousand times, tabulating the number of times the simulated data leads to a rejection of the null hypothesis. The power is the proportion of simulated samples in step 2 that lead to rejection. The significance level is the proportion of simulated samples in step 3 that lead to rejection.

Generating Random Distributions

Two methods are available in *PASS* to simulate random samples. The first method generates the random variates directly, one value at a time. The second method generates a large pool (over 10,000) of random values and then draw the random numbers from this pool. This second method can cut the running time of the simulation by 70%.

As mentioned above, the second method begins by generating a large pool of random numbers from the specified distributions. Each of these pools is evaluated to determine if its mean is within a small relative tolerance (0.0001) of the target mean. If the actual mean is not within the tolerance of the target mean, individual members of the population are replaced with new random numbers if the new random number moves the mean towards its target. Only a few hundred such swaps are required to bring the actual mean to within tolerance of the target mean. This population is then sampled with replacement using the uniform distribution. We have found that this method works well as long as the size of the pool is the maximum of twice the number of simulated samples desired and 10,000.

Simulating Data for an Equivalence Test

Simulating equivalence data is more complex than simulating data for a regular two-sided test. An equivalence test essentially reverses the roles of the null and alternative hypothesis. The null hypothesis becomes

$$H_0: (\mu_1 - \mu_2) \leq -D \text{ or } (\mu_1 - \mu_2) \geq D$$

where D is the margin of equivalence. Thus the null hypothesis is made up of two simple hypotheses:

$$H_{0_1}: (\mu_1 - \mu_2) \leq -D$$

$$H_{0_2}: (\mu_1 - \mu_2) \geq D$$

The additional complexity comes in deciding which of the two null hypotheses are used to simulate data for the null hypothesis situation. The choice becomes more problematic when asymmetric equivalence limits are chosen. In this case, you may want to try simulating using each simple null hypothesis in turn.

To generate data for the null hypotheses, generate data for each group. The difference in the means of these two groups will become one of the equivalence limits. The other equivalence limit will be determined by symmetry and will always have a sign that is the opposite of the first equivalence limit.

Test Statistics

This section describes the test statistics that are available in this procedure. Note that these test statistics are computed on the differences. Thus, when the equation refers to an X value, this X value is assumed to be a difference between two individual variates.

Equivalence Tests for Two Means (Simulation)

Two-Sample T-Test

The t-test assumes that the data are simple random samples from populations of normally-distributed values that have the same mean and variance. This assumption implies that the data are continuous and their distribution is symmetric. The calculation of the t statistic is as follows.

$$t_{df} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}}$$

where

$$\bar{X}_k = \frac{\sum_{i=1}^{N_k} X_{ki}}{N_k}$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_1 + N_2 - 2} \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}$$

$$df = N_1 + N_2 - 2$$

The significance of the test statistic is determined by computing a p-value which is based on the t distribution with appropriate degrees of freedom. If this p-value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.

Welch's T-Test

Welch (1938) proposed the following test for use when the two variances are not assumed to be equal.

$$t_f^* = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}^*}$$

where

$$s_{\bar{X}_1 - \bar{X}_2}^* = \sqrt{\left(\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2}{N_1(N_1 - 1)} \right) + \left(\frac{\sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_2(N_2 - 1)} \right)}$$

$$f = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} \right)^2}{\frac{s_1^4}{N_1^2(N_1 - 1)} + \frac{s_2^4}{N_2^2(N_2 - 1)}}$$

$$s_1 = \sqrt{\left(\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2}{N_1 - 1} \right)}, \quad s_2 = \sqrt{\left(\frac{\sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_2 - 1} \right)}$$

Equivalence Tests for Two Means (Simulation)

Trimmed T-Test assuming Equal Variances

The notion of trimming off a small proportion of possibly outlying observations and using the remaining data to form a t-test was first proposed for one sample by Tukey and McLaughlin (1963). Dixon and Tukey (1968) consider a slight modification of this test, called *Winsorization*, which replaces the trimmed data with the nearest remaining value. The two-sample trimmed t-test was proposed by Yuen and Dixon (1973).

Assume that the data values have been sorted from lowest to highest. The *trimmed mean* is defined as

$$\bar{X}_{tg} = \frac{\sum_{k=g+1}^{N-g} X_k}{h}$$

where $h = N - 2g$ and $g = [N(G/100)]$. Here we use $[Z]$ to mean the largest integer smaller than Z with the modification that if G is non-zero, the value of $[N(G/100)]$ is at least one. G is the percent trimming and should usually be less than 25%, often between 5% and 10%. Thus, the g smallest and g largest observation are omitted in the calculation.

To calculate the modified t-test, calculate the *Winsorized mean* and the *Winsorized* sum of squared deviations as follows.

$$\bar{X}_{wg} = \frac{g(X_{g+1} + X_{N-g}) + \sum_{k=g+1}^{N-g} X_k}{N}$$

$$SSD_{wg} = \frac{g(X_{g+1} - \bar{X}_{wg})^2 + g(X_{N-g} - \bar{X}_{wg})^2 + \sum_{k=g+1}^{N-g} (X_k - \bar{X}_{wg})^2}{N}$$

Using the above definitions, the two-sample trimmed t-test is given by

$$T_{tg} = \frac{(\bar{X}_{1tg} - \bar{X}_{2tg}) - (\mu_1 - \mu_2)}{\sqrt{\frac{SSD_{1wg} + SSD_{2wg}}{h_1 + h_2 - 2} \left(\frac{1}{h_1} + \frac{1}{h_2} \right)}}$$

The distribution of this t statistic is approximately that of a t distribution with degrees of freedom equal to $h_1 + h_2 - 2$. This approximation is often reasonably accurate if both sample sizes are greater than 6.

Trimmed T-Test assuming Unequal Variances

Yuen (1974) combines trimming (see above) with Welch's (1938) test. The resulting trimmed Welch test is resistant to outliers and seems to alleviate some of the problems that occur because of skewness in the underlying distributions. Extending the results from above, the trimmed version of Welch's t-test is given by

$$T_{tg}^* = \frac{(\bar{X}_{1tg} - \bar{X}_{2tg}) - (\mu_1 - \mu_2)}{\sqrt{\frac{SSD_{1wg}}{h_1(h_1 - 1)} + \frac{SSD_{2wg}}{h_2(h_2 - 1)}}}$$

with degrees of freedom f given by

$$\frac{1}{f} = \frac{c^2}{h_1 - 1} + \frac{1 - c^2}{h_2 - 1}$$

Equivalence Tests for Two Means (Simulation)

where

$$c = \frac{\frac{SSD_{1wg}}{h_1(h_1 - 1)}}{\frac{SSD_{1wg}}{h_1(h_1 - 1)} + \frac{SSD_{2wg}}{h_2(h_2 - 1)}}$$

Mann-Whitney U Test

This test is the nonparametric substitute for the equal-variance t-test. Two key assumptions for this test are that the distributions are at least ordinal and that they are identical under H0. This means that ties (repeated values) are not acceptable. When ties are present, an approximation can be used, but the theoretic results no longer hold.

The Mann-Whitney test statistic is defined as follows in Gibbons (1985).

$$z = \frac{W_1 - \frac{N_1(N_1 + N_2 + 1)}{2} + C}{s_w}$$

where

$$W_1 = \sum_{k=1}^{N_1} Rank(X_{1k})$$

The ranks are determined after combining the two samples. The standard deviation is calculated as

$$s_w = \sqrt{\frac{N_1 N_2 (N_1 + N_2 + 1)}{12} - \frac{N_1 N_2 \sum_{i=1} (t_i^3 - t_i)}{12(N_1 + N_2)(N_1 + N_2 - 1)}}$$

where t_i is the number of observations tied at value one, t_2 is the number of observations tied at some value two, and so forth.

The correction factor, C , is 0.5 if the rest of the numerator of z is negative or -0.5 otherwise. The value of z is then compared to the standard normal distribution.

Standard Deviations

Care must be used when either the null or alternative distribution is not normal. In these cases, the standard deviation is usually not specified directly. For example, you might use a gamma distribution with a shape parameter of 1.5 and a mean of 4 as the null distribution and a gamma distribution with the same shape parameter and a mean of 5 as the alternative distribution. This allows you to compare the two means. However, although the shape parameters are constant, the standard deviations, which are based on both the shape parameter and the mean, are not. Thus the distributions not only have different means, but different standard deviations!

Equivalence Tests for Two Means (Simulation)

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be calculated using the values of the other parameters. Under most conditions, you would select either *Power* or *Sample Size (NI)*.

Select *Power* when you want to estimate the power for a specific scenario.

Select *Sample Size (NI)* when you want to determine the sample size needed to achieve a given power and alpha level. This option is computationally intensive and may take a long time to complete.

Test and Simulations

Equivalence Limit

Equivalence limits are defined as the positive and negative limits around zero that define a zone of equivalence. This zone of equivalence is a set of difference values that define a region in which the two means are 'close enough' so that they are considered to be the same for practical purposes.

Rather than define these limits explicitly, they are set implicitly. This is done as follows. One limit is found by subtracting the Group 2 Dist'n|H0 mean from the Group 1 Dist'n|H0 mean. If the limits are symmetric, the other limit is this difference times -1. To obtain symmetric limits, enter 'Symmetric' here.

If asymmetric limits are desired, a numerical value is specified here. It is given the sign (+ or -) that is opposite the difference of the means discussed above.

For example, if the mean of group 1 under H0 is 5, the mean of group 2 under H0 is 4, and *Symmetric* is entered here, the equivalence limits will be $5 - 4 = 1$ and -1 . However, if the value *1.25* is entered here, the equivalence limits are 1 and -1.25 .

If you do not have a specific value in mind for the equivalence limit, a common value for an equivalence limit is 20% or 25% of the group 1 (reference) mean.

Test Type

Specify which test statistic is to be used in the simulation. Although the t-test is the most commonly used test statistic, it is based on assumptions that may not be viable in many situations. For your data, you may find that one of the other tests are more accurate (actual alpha = target alpha) and more precise (better power).

Simulations

This option specifies the number of iterations, M , used in the simulation. The larger the number of iterations, the longer the running time, and, the more accurate the results.

The precision of the simulated power estimates are calculated from the binomial distribution. Thus, confidence intervals may be constructed for various power values. The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is either 0.50 or 0.95. The table values are interpreted as follows: a 95% confidence interval of the true power is given by the power reported by the simulation plus and minus the 'Precision' amount given in the table.

Equivalence Tests for Two Means (Simulation)

Simulation Size M	Precision when Power = 0.50	Precision when Power = 0.95
100	0.100	0.044
500	0.045	0.019
1000	0.032	0.014
2000	0.022	0.010
5000	0.014	0.006
10000	0.010	0.004
50000	0.004	0.002
100000	0.003	0.001

Notice that a simulation size of 1000 gives a precision of plus or minus 0.01 when the true power is 0.95. Also note that as the simulation size is increased beyond 5000, there is only a small amount of additional accuracy achieved.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of nonequivalent means when in fact the means are equivalent.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of non-equivalent means when in fact the means are nonequivalent.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Sample Size (When Solving for Sample Size)

Group Allocation

Select the option that describes the constraints on $N1$ or $N2$ or both.

The options are

- **Equal ($N1 = N2$)**

This selection is used when you wish to have equal sample sizes in each group. Since you are solving for both sample sizes at once, no additional sample size parameters need to be entered.

- **Enter $N2$, solve for $N1$**

Select this option when you wish to fix $N2$ at some value (or values), and then solve only for $N1$. Please note that for some values of $N2$, there may not be a value of $N1$ that is large enough to obtain the desired power.

Equivalence Tests for Two Means (Simulation)

- **Enter $R = N2/N1$, solve for $N1$ and $N2$**

For this choice, you set a value for the ratio of $N2$ to $N1$, and then PASS determines the needed $N1$ and $N2$, with this ratio, to obtain the desired power. An equivalent representation of the ratio, R , is

$$N2 = R * N1.$$

- **Enter percentage in Group 1, solve for $N1$ and $N2$**

For this choice, you set a value for the percentage of the total sample size that is in Group 1, and then PASS determines the needed $N1$ and $N2$ with this percentage to obtain the desired power.

$N2$ (Sample Size, Group 2)

This option is displayed if Group Allocation = "Enter $N2$, solve for $N1$ "

$N2$ is the number of items or individuals sampled from the Group 2 population.

$N2$ must be ≥ 2 . You can enter a single value or a series of values.

R (Group Sample Size Ratio)

This option is displayed only if Group Allocation = "Enter $R = N2/N1$, solve for $N1$ and $N2$."

R is the ratio of $N2$ to $N1$. That is,

$$R = N2 / N1.$$

Use this value to fix the ratio of $N2$ to $N1$ while solving for $N1$ and $N2$. Only sample size combinations with this ratio are considered.

$N2$ is related to $N1$ by the formula:

$$N2 = [R \times N1],$$

where the value $[Y]$ is the next integer $\geq Y$.

For example, setting $R = 2.0$ results in a Group 2 sample size that is double the sample size in Group 1 (e.g., $N1 = 10$ and $N2 = 20$, or $N1 = 50$ and $N2 = 100$).

R must be greater than 0. If $R < 1$, then $N2$ will be less than $N1$; if $R > 1$, then $N2$ will be greater than $N1$. You can enter a single or a series of values.

Percent in Group 1

This option is displayed only if Group Allocation = "Enter percentage in Group 1, solve for $N1$ and $N2$."

Use this value to fix the percentage of the total sample size allocated to Group 1 while solving for $N1$ and $N2$. Only sample size combinations with this Group 1 percentage are considered. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single or a series of values.

Equivalence Tests for Two Means (Simulation)

Sample Size (When Not Solving for Sample Size)

Group Allocation

Select the option that describes how individuals in the study will be allocated to Group 1 and to Group 2.

The options are

- **Equal ($N1 = N2$)**
This selection is used when you wish to have equal sample sizes in each group. A single per group sample size will be entered.
- **Enter $N1$ and $N2$ individually**
This choice permits you to enter different values for $N1$ and $N2$.
- **Enter $N1$ and R , where $N2 = R * N1$**
Choose this option to specify a value (or values) for $N1$, and obtain $N2$ as a ratio (multiple) of $N1$.
- **Enter total sample size and percentage in Group 1**
Choose this option to specify a value (or values) for the total sample size (N), obtain $N1$ as a percentage of N , and then $N2$ as $N - N1$.

Sample Size Per Group

This option is displayed only if Group Allocation = "Equal ($N1 = N2$)."

The Sample Size Per Group is the number of items or individuals sampled from each of the Group 1 and Group 2 populations. Since the sample sizes are the same in each group, this value is the value for $N1$, and also the value for $N2$.

The Sample Size Per Group must be ≥ 2 . You can enter a single value or a series of values.

$N1$ (Sample Size, Group 1)

*This option is displayed if Group Allocation = "Enter $N1$ and $N2$ individually" or "Enter $N1$ and R , where $N2 = R * N1$."*

$N1$ is the number of items or individuals sampled from the Group 1 population.

$N1$ must be ≥ 2 . You can enter a single value or a series of values.

$N2$ (Sample Size, Group 2)

This option is displayed only if Group Allocation = "Enter $N1$ and $N2$ individually."

$N2$ is the number of items or individuals sampled from the Group 2 population.

$N2$ must be ≥ 2 . You can enter a single value or a series of values.

R (Group Sample Size Ratio)

*This option is displayed only if Group Allocation = "Enter $N1$ and R , where $N2 = R * N1$."*

R is the ratio of $N2$ to $N1$. That is,

$$R = N2/N1$$

Use this value to obtain $N2$ as a multiple (or proportion) of $N1$.

$N2$ is calculated from $N1$ using the formula:

$$N2 = [R \times N1],$$

where the value $[Y]$ is the next integer $\geq Y$.

Equivalence Tests for Two Means (Simulation)

For example, setting $R = 2.0$ results in a Group 2 sample size that is double the sample size in Group 1.

R must be greater than 0. If $R < 1$, then $N2$ will be less than $N1$; if $R > 1$, then $N2$ will be greater than $N1$. You can enter a single value or a series of values.

Total Sample Size (N)

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This is the total sample size, or the sum of the two group sample sizes. This value, along with the percentage of the total sample size in Group 1, implicitly defines $N1$ and $N2$.

The total sample size must be greater than one, but practically, must be greater than 3, since each group sample size needs to be at least 2.

You can enter a single value or a series of values.

Percent in Group 1

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This value fixes the percentage of the total sample size allocated to Group 1. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single value or a series of values.

Effect Size

Group 1 & 2 Distributions|H0 (Null)

These options specify the distributions of the two groups under the null hypothesis, H_0 . The difference between the means of these two distributions is the value of one of the equivalence limits.

Group 1 is often called the reference (or standard) distribution. Group 2 is often called the treatment distribution. These options specify these two distributions under the null hypothesis, H_0 . The difference between the means of these two distributions is, by definition, one of the equivalence limits. Thus, you set the equivalence limit by specifying the two means.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The characters $M0$ and $M1$ are to be used for the means of the distributions of groups 1 and 2 under H_0 , respectively. An equivalence limit is then $M0 - M1$, which must be non-zero.

For example, suppose you entered $Normal(M0 S)$ for group 1 and $Normal(M1 S)$ for group 2. Also, you set $M0$ equal to 5 and $M1$ equal to 4. The upper (positive) equivalence limit would be $5 - 4 = 1$.

Following is a list of the distributions that are available and the syntax used to specify them. Each of the parameters should be replaced with a number or parameter name.

Distributions with Common Parameters

Beta(Shape1, Shape2, Min, Max)

Binomial(P, N)

Cauchy(Mean, Scale)

Constant(Value)

Exponential(Mean)

Gamma(Shape, Scale)

Gumbel(Location, Scale)

Equivalence Tests for Two Means (Simulation)

Laplace(Location, Scale)
 Logistic(Location, Scale)
 Lognormal(Mu, Sigma)
 Multinomial(P1, P2, P3, ..., Pk)
 Normal(Mean, Sigma)
 Poisson(Mean)
 TukeyGH(Mu, S, G, H)
 Uniform(Min, Max)
 Weibull(Shape, Scale)

Distributions with Mean and SD Parameters

BetaMS(Mean, SD, Min, Max)
 BinomialMS(Mean, N)
 GammaMS(Mean, SD)
 GumbelMS(Mean, SD)
 LaplaceMS(Mean, SD)
 LogisticMS(Mean, SD)
 LognormalMS(Mean, SD)
 UniformMS(Mean, SD)
 WeibullMS(Mean, SD)

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on *Data Simulation* and will not be repeated here.

Finding the Value of the Mean of a Specified Distribution

Most of the distributions have been parameterized in terms of their means, since this is the parameter being tested. The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a distribution created by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of $4N(4, 5) + 2N(5, 6)$ is $4*4 + 2*5 = 26$, but the mean of $4N(4, 5) * 2N(5, 6)$ is not exactly $4*4*2*5 = 160$ (although it is close).

Group 1 & 2 Distributions|H1 (Alternative)

These options specify the distributions of the two groups under the alternative hypothesis, H1. The difference between the means of these two distributions is the difference that is assumed to be the true value of the difference.

Usually, the mean difference is specified by entering MO for the mean parameter in the distribution expression for group 1 and MI for the mean parameter in the distribution expression for group 2. The mean difference under H1 then becomes the value of $MO - MI = 0$. If you want a non-zero value, you specify it by specifying unequal values for the two distribution means. For example, you could enter A for the mean of group 2. The mean difference will then be $MO - A$.

Equivalence Tests for Two Means (Simulation)

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value $M1$ is reserved for the value of the mean of group 2 under the alternative hypothesis.

A list of the distributions that are available and the syntax used to specify them is given above.

Effect Size – Distribution Parameters

M0 (Mean|H0)

These values are substituted for $M0$ in the distribution specifications given above. $M0$ is intended to be the value of the mean hypothesized by the null hypothesis, $H0$.

You can enter a list of values using the syntax $0\ 1\ 2\ 3$ or $0\ to\ 3\ by\ 1$.

M1 (Mean|H1)

These values are substituted for $M1$ in the distribution specifications given above. Although it can be used wherever you want, $M1$ is intended to be the value of the mean hypothesized by the alternative hypothesis, $H1$.

You can enter a list of values using the syntax $0\ 1\ 2\ 3$ or $0\ to\ 3\ by\ 1$.

Parameter Values (S, A, B)

Enter the numeric value(s) of parameter listed above. These values are substituted for the corresponding letter in all four distribution specifications.

You can enter a list of values using the syntax '0 2 3' or '0 to 3 by 1.'

You can also change the letter than is used as the name of this parameter.

Options Tab

The Options tab contains calculation options.

Random Numbers

Random Number Pool Size

This is the size of the pool of values from which the random samples will be drawn. Pools should be at least the maximum of 10,000 and twice the number of simulations. You can enter *Automatic* and an appropriate value will be calculated.

If you do not want to draw numbers from a pool, enter 0 here.

Trimmed T-Test

Percent Trimmed at Each End

Specify the percent of each end of the sorted data that is to be trimmed (constant G above) when using the trimmed means procedures. This percentage is applied to the sample size to determine how many of the lowest and highest data values are to be trimmed by the procedure. For example, if the sample size ($N1$) is 27 and you specify 10 here, then $[27*10/100] = 2$ observations will be trimmed at the bottom and the top. For any percentage, at least one observation is trimmed from each end of the sorted dataset.

The range of possible values is 0 to 25.

Equivalence Tests for Two Means (Simulation)

Example 1 – Power at Various Sample Sizes

Researchers are planning an experiment to determine if the response to a new drug is equivalent to the response to the standard drug. The average response level to the standard drug is known to be 63 with a standard deviation of 5. The researchers decide that if the average response level to the new drug is between 60 and 66, they will consider it to be equivalent to the standard drug.

The researchers decide to use a parallel-group design. The response level for the standard drug will be measured for each subject. They will analyze the data using an equivalence test based on the t-test with an alpha level of 0.05. They want to compare the power at sample sizes of 10, 30, 50, and 70. They assume that the data are normally distributed and that the true difference between the mean response of the two drugs is zero. Since this is an exploratory analysis, the number of simulation iterations is set to 2000.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Equivalence Limit.....	Symmetric
Test Type.....	T-Test
Simulations	2000
Alpha.....	0.05
Group Allocation	Equal (N1 = N2)
Sample Size Per Group.....	10 30 50 70
Group 1 Distribution H0	Normal(M0 S)
Group 2 Distribution H0	Normal(M1 S)
Group 1 Distribution H1	Normal(M0 S)
Group 2 Distribution H1	Normal(M0 S)
M0 (Mean H0).....	63
M1 (Mean H1).....	66
S.....	5

Equivalence Tests for Two Means (Simulation)

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results and Plots

Numeric Results for Testing Mean Equivalence. Hypotheses: $H_0: \text{Diff} \geq |\text{Diff}_0|$; $H_1: \text{Diff} < |\text{Diff}_0|$
 H0 Dist's: Normal(M0 S) & Normal(M1 S)
 H1 Dist's: Normal(M0 S) & Normal(M0 S)
 Test Statistic: T-Test

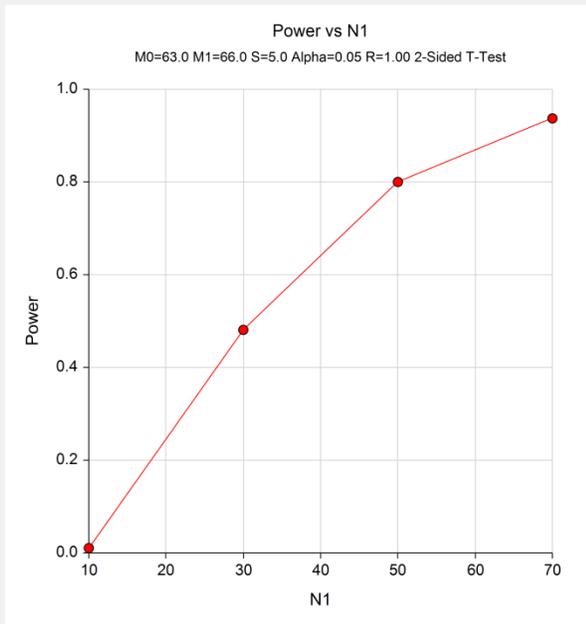
Power	N1/N2	H1 Diff1	Lower Equiv. Limit	Upper Equiv. Limit	Target Alpha	Actual Alpha	Beta	M0	M1	S
0.009 (0.004)	10/10 [0.005]	0.0 0.013]	-3.0	3.0	0.050	0.005 (0.003)	0.991 [0.002]	63.0 0.008]	66.0	5.0
0.477 (0.022)	30/30 [0.455]	0.0 0.498]	-3.0	3.0	0.050	0.053 (0.010)	0.524 [0.043]	63.0 0.062]	66.0	5.0
0.816 (0.017)	50/50 [0.799]	0.0 0.833]	-3.0	3.0	0.050	0.061 (0.010)	0.184 [0.050]	63.0 0.071]	66.0	5.0
0.944 (0.010)	70/70 [0.934]	0.0 0.954]	-3.0	3.0	0.050	0.050 (0.010)	0.056 [0.040]	63.0 0.060]	66.0	5.0

Notes:
 Pool Size: 10000. Simulations: 2000. Run Time: 21.61 seconds.

Summary Statements

Group sample sizes of 10 and 10 achieve 1% power to detect equivalence when the margin of equivalence is from -3.0 to 3.0 and the actual mean difference is 0.0. The significance level (alpha) is 0.050 using two one-sided T-Tests. These results are based on 2000 Monte Carlo samples from the null distributions: Normal(M0 S) and Normal(M1 S), and the alternative distributions: Normal(M0 S) and Normal(M0 S).

Chart Section



This report shows the estimated power for each scenario. The first row shows the parameter settings and the estimated power and significance level (Actual Alpha). The second row shows two 95% confidence intervals in brackets: the first for the power and the second for the significance level. Half the width of each confidence interval is given in parentheses as a fundamental measure of the accuracy of the simulation. As the number of simulations is increased, the width of the confidence intervals will decrease.

Equivalence Tests for Two Means (Simulation)

Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to determine how large a sample is needed to obtain a power of 0.90.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size (N1)
Equivalence Limit.....	Symmetric
Test Type.....	T-Test
Simulations	2000
Power.....	0.90
Alpha.....	0.05
Group Allocation	Equal (N1 = N2)
Group 1 Distribution H0	Normal(M0 S)
Group 2 Distribution H0	Normal(M1 S)
Group 1 Distribution H1	Normal(M0 S)
Group 2 Distribution H1	Normal(M0 S)
M0 (Mean H0).....	63
M1 (Mean H1).....	66
S.....	5

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Testing Mean Equivalence. Hypotheses: H0: Diff \geq Diff0 ; H1: Diff < Diff0										
H0 Dist's: Normal(M0 S) & Normal(M1 S)										
H1 Dist's: Normal(M0 S) & Normal(M0 S)										
Test Statistic: T-Test										
Power	N1/N2	H1 Diff1	Lower Equiv. Limit	Upper Equiv. Limit	Target Alpha	Actual Alpha	Beta	M0	M1	S
0.911 (0.012)	61/61 [0.899]	0.0 0.923]	-3.0	3.0	0.050	0.044 (0.009)	0.089 [0.035]	63.0 0.053]	66.0	5.0

The required sample size is 61 per group.

Equivalence Tests for Two Means (Simulation)

Example 3 – Comparative Results when the Data Contain Outliers

Continuing Example 1, this example will investigate the impact of outliers on the characteristics of the various test statistics. The two-sample t-test is known to be robust to the violation of some assumptions, but it is susceptible to inaccuracy when the data contains outliers. This example will investigate the impact of outliers on the power and precision of the five test statistics available in *PASS*.

A mixture of two normal distributions will be used to randomly generate outliers. The mixture will draw 95% of the data from a standard distribution. The other 5% of the data will come from a normal distribution with the same mean but with a standard deviation that is one, five, and ten times larger than that of the standard.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the *PASS* Home window, load the **Equivalence Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Equivalence Limit.....	Symmetric
Test Type.....	T-Test
Simulations	2000
Alpha.....	0.05
Group Allocation	Equal (N1 = N2)
Sample Size Per Group.....	40
Group 1 Distribution H0	Normal(M0 S)[95];Normal(M0 A)[5]
Group 2 Distribution H0	Normal(M1 S)[95];Normal(M1 A)[5]
Group 1 Distribution H1	Normal(M0 S)[95];Normal(M0 A)[5]
Group 2 Distribution H1	Normal(M0 S)[95];Normal(M0 A)[5]
M0 (Mean H0)	63
M1 (Mean H1)	66
S.....	5
A.....	5 25 50
Reports Tab	
Show Comparative Reports.....	Checked
Plots Tab	
Show Comparative Plots	Checked

Equivalence Tests for Two Means (Simulation)

Output

Click the Calculate button to perform the calculations and generate the following output.

Power Comparison for Testing Equivalence. Hypotheses: $H_0: \text{Diff} \geq |\text{Diff}_0|$; $H_1: \text{Diff} < |\text{Diff}_0|$
 H0 Dist's: Normal(M0 S)[95];Normal(M0 A)[5] & Normal(M1 S)[95];Normal(M1 A)[5]
 H1 Dist's: Normal(M0 S)[95];Normal(M0 A)[5] & Normal(M0 S)[95];Normal(M0 A)[5]

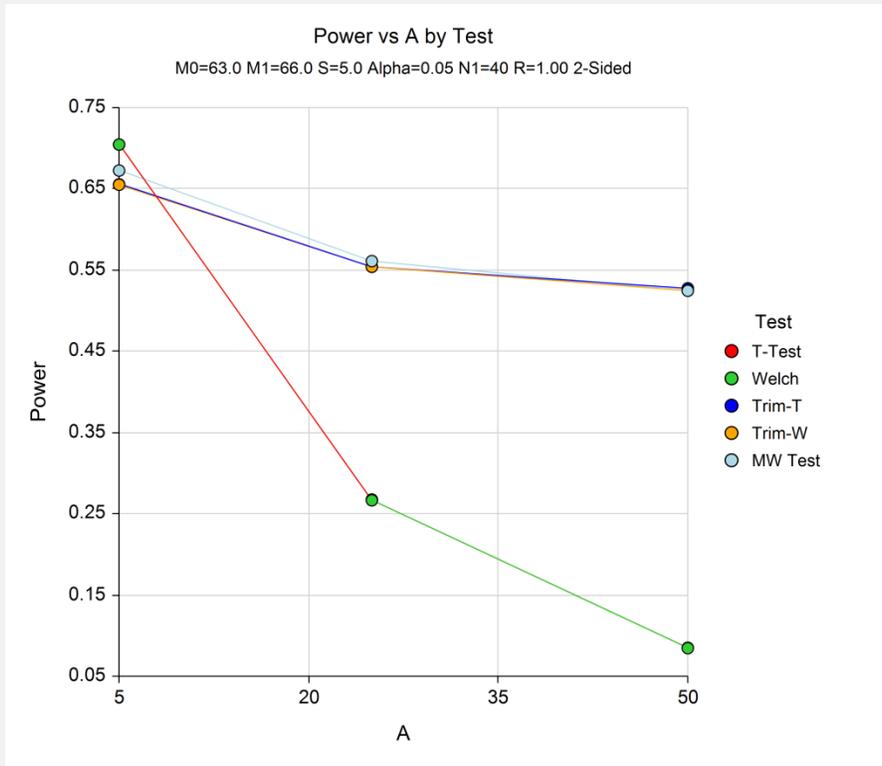
N1/N2 (Diff1)	H1 Diff	Lower Equiv. Limit	Upper Equiv. Limit	Target Alpha	T-Test Power	Welch Power	Trim. T-Test Power	Trim. Welch Power	Mann Whit'y Power	M0	M1	S	A
40/40	0.0	-3.0	3.0	0.050	0.708	0.708	0.657	0.656	0.672	63.0	66.0	5.0	5.0
40/40	0.0	-3.0	3.0	0.050	0.247	0.247	0.543	0.543	0.539	63.0	66.0	5.0	25.0
40/40	0.0	-3.0	3.0	0.050	0.073	0.072	0.509	0.508	0.510	63.0	66.0	5.0	50.0

Pool Size: 10000. Simulations: 2000. Run Time: 2.90 minutes. Percent Trimmed: 10.

Alpha Comparison for Testing Equivalence. Hypotheses: $H_0: \text{Diff} \geq |\text{Diff}_0|$; $H_1: \text{Diff} < |\text{Diff}_0|$
 H0 Dist's: Normal(M0 S)[95];Normal(M0 A)[5] & Normal(M1 S)[95];Normal(M1 A)[5]
 H1 Dist's: Normal(M0 S)[95];Normal(M0 A)[5] & Normal(M0 S)[95];Normal(M0 A)[5]

N1/N2 (Diff1)	H1 Diff	Lower Equiv. Limit	Upper Equiv. Limit	Target Alpha	T-Test Alpha	Welch Alpha	Trim. T-Test Alpha	Trim. Welch Alpha	Mann Whit'y Alpha	M0	M1	S	A
40/40	0.0	-3.0	3.0	0.050	0.050	0.050	0.058	0.058	0.056	63.0	66.0	5.0	5.0
40/40	0.0	-3.0	3.0	0.050	0.030	0.030	0.041	0.041	0.044	63.0	66.0	5.0	25.0
40/40	0.0	-3.0	3.0	0.050	0.008	0.008	0.042	0.042	0.044	63.0	66.0	5.0	50.0

Pool Size: 10000. Simulations: 2000. Run Time: 2.90 minutes. Percent Trimmed: 10.



When A = 5, there are no outliers and the power of the nonparametric test and the trimmed tests are a little less than that of the t-test. When A = 25, the distortion of the t-test caused by the outliers becomes apparent. In this case, the powers of the standard t-test and Welch’s t-test are 0.247, but the powers of the nonparametric Mann-Whitney test and the trimmed tests are about 0.54. When A = 50, the standard t-test only achieves a power of 0.073, but the trimmed and nonparametric tests achieve powers of about 0.51!

Looking at the second table, we see that the true significance level of the t-test is distorted by the outliers, while the significance levels of the other tests remain close to the target value.

Equivalence Tests for Two Means (Simulation)

Example 4 – Selecting a Test Statistic when the Data Are Skewed

Continuing Example 3, this example will investigate the impact of skewness in the underlying distribution on the characteristics of the various test statistics.

Tukey's G-H distribution will be used because it allows the amount of skewness to be gradually increased.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Equivalence Limit.....	Symmetric
Test Type.....	T-Test
Simulations	2000
Alpha.....	0.05
Group Allocation	Equal (N1 = N2)
Sample Size Per Group	40
Group 1 Distribution H0	TukeyGH(M0 S G 0)
Group 2 Distribution H0	TukeyGH(M1 S G 0)
Group 1 Distribution H1	TukeyGH(M0 S G 0)
Group 2 Distribution H1	TukeyGH(M0 S G 0)
M0 (Mean H0).....	63
M1 (Mean H1).....	66
S.....	5
G	0 0.5 0.9 (note that A has been changed to G)
Reports Tab	
Show Comparative Reports.....	Checked
Plots Tab	
Show Comparative Plots	Checked

Equivalence Tests for Two Means (Simulation)

Output

Click the Calculate button to perform the calculations and generate the following output.

Power Comparison for Testing Equivalence. Hypotheses: $H_0: \text{Diff} \geq |\text{Diff}_0|$; $H_1: \text{Diff} < |\text{Diff}_0|$
H0 Dist's: TukeyGH(M0 S G 0) & TukeyGH(M1 S G 0)
H1 Dist's: TukeyGH(M0 S G 0) & TukeyGH(M0 S G 0)

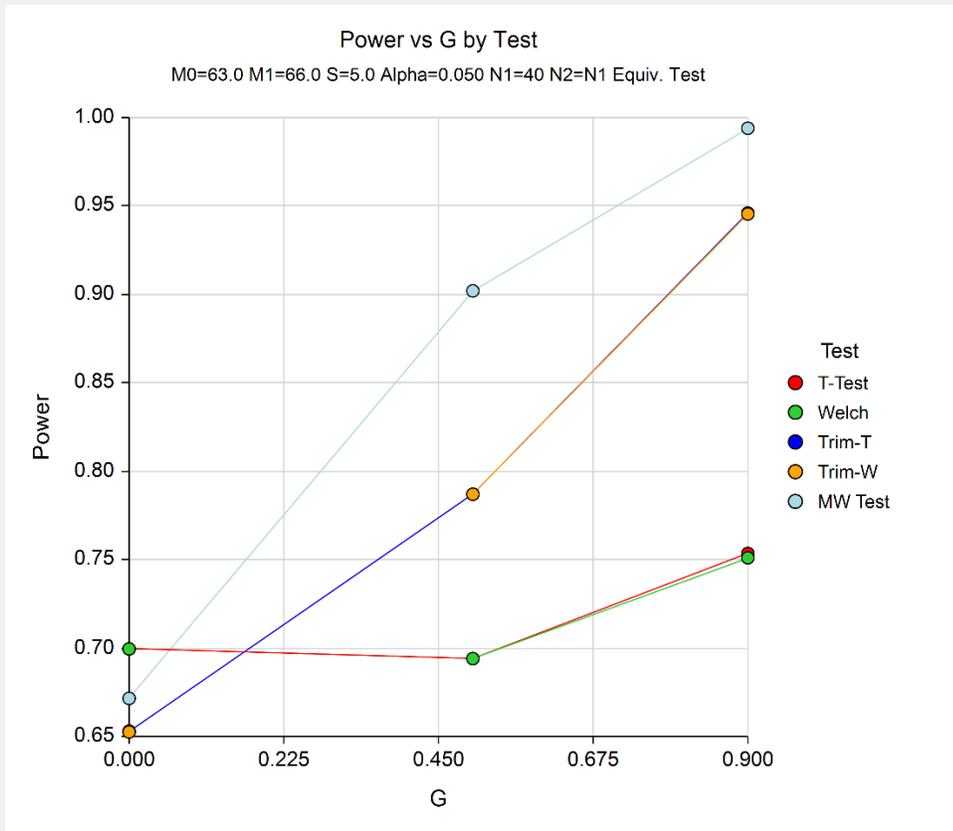
N1/N2 (Diff1)	H1 Diff	Lower Equiv. Limit	Upper Equiv. Limit	Target Alpha	T-Test Power	Trim.		Mann Whit'y Power	M0	M1	S	G
						Welch Power	T-Test Power					
40/40	0.0	-3.0	3.0	0.050	0.700	0.700	0.653	0.653	63.0	66.0	5.0	0.0
40/40	0.0	-3.0	3.0	0.050	0.694	0.694	0.787	0.787	63.0	66.0	5.0	0.5
40/40	0.0	-3.0	3.0	0.050	0.754	0.751	0.946	0.946	63.0	66.0	5.0	0.9

Pool Size: 10000. Simulations: 2000. Run Time: 23.54 seconds. Percent Trimmed at each end: 10.

Alpha Comparison for Testing Equivalence. Hypotheses: $H_0: \text{Diff} \geq |\text{Diff}_0|$; $H_1: \text{Diff} < |\text{Diff}_0|$
H0 Dist's: TukeyGH(M0 S G 0) & TukeyGH(M1 S G 0)
H1 Dist's: TukeyGH(M0 S G 0) & TukeyGH(M0 S G 0)

N1/N2 (Diff1)	H1 Diff	Lower Equiv. Limit	Upper Equiv. Limit	Target Alpha	T-Test Alpha	Trim.		Mann Whit'y Alpha	M0	M1	S	G
						Welch Alpha	T-Test Alpha					
40/40	0.0	-3.0	3.0	0.050	0.053	0.053	0.053	0.053	63.0	66.0	5.0	0.0
40/40	0.0	-3.0	3.0	0.050	0.059	0.059	0.050	0.052	63.0	66.0	5.0	0.5
40/40	0.0	-3.0	3.0	0.050	0.047	0.046	0.045	0.051	63.0	66.0	5.0	0.9

Pool Size: 10000. Simulations: 2000. Run Time: 23.54 seconds. Percent Trimmed at each end: 10.



We see that as the degree of skewness is increased, the power of the t-test increases slightly, but the powers of the trimmed and nonparametric tests improve dramatically. The significance levels do not appear to be adversely impacted.

Equivalence Tests for Two Means (Simulation)

Example 5 – Validation using Machin

Machin *et al.* (1997) page 107 present an example of determining the sample size for a parallel-group design in which the reference mean is 96, the treatment mean is 94, the standard deviation is 8, the limits are plus or minus 5, the power is 80%, and the significance level is 0.05. They calculate the sample size to be 88. It is important to note that Machin *et al.* use an approximation, so their results cannot be expected to exactly match those of *PASS*.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Equivalence Limit.....	Symmetric
Test Type.....	T-Test
Simulations	2000
Power.....	0.80
Alpha.....	0.05
Group Allocation	Equal (N1 = N2)
Group 1 Distribution H0	Normal(M0 S)
Group 2 Distribution H0	Normal(91 S)
Group 1 Distribution H1	Normal(M0 S)
Group 2 Distribution H1	Normal(94 S)
M0 (Mean H0)	96
M1 (Mean H1)	1
S.....	8

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results for Testing Mean Equivalence. Hypotheses: H0: Diff \geq |Diff0|; H1: Diff < |Diff0|
H0 Dist's: Normal(M0 S) & Normal(91 S)
H1 Dist's: Normal(M0 S) & Normal(94 S)
Test Statistic: T-Test

Power	N1/N2	H1 Diff1	Lower Equiv. Limit	Upper Equiv. Limit	Target Alpha	Actual Alpha	Beta	M0	S
0.807 (0.017)	87/87 [0.790]	2.0 0.824]	-5.0	5.0	0.050	0.049 (0.009)	0.193 [0.039]	96.0 0.058]	8.0

Notes:

Pool Size: 10000. Simulations: 2000. Run Time: 60.05 seconds.

The sample size of 87 per group is reasonably close to the analytic answer of 88.